OM protein - protein search, using sw model

Run on: May 29, 2003, 15:21:44 ; Search time 22 Seconds
(without alignments)
152.708 Million cell updates/sec

Title: US-09-924-102-2
Perfect score: 81
Sequence: 1 MLLSTHLFIYLFIYFLSYSL.....RWGGQGGRGGTADTGGMFLS 81

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8	9.9	65	1	RPR_DROME	Q24475 drosophila
2	8	9.9	284	1	SUHA_HUMAN	Q06520 homo sapien
3	8	9.9	284	1	SUHA_MACFA	P52842 macaca fasc
4	8	9.9	587	1	ALU2_HUMAN	P39189 homo sapien
5	8	9.9	587	1	ALU3_HUMAN	P39190 homo sapien
6	8	9.9	591	1	ALU8_HUMAN	P39195 homo sapien
7	7	8.6	116	1	FLHD_XENNE	Q9x9f2 xenorhabdus
8	7	8.6	285	1	INVO_CANFA	P18174 canis famil
9	7	8.6	304	1	Y687_HAEIN	P71356 haemophilus
10	7	8.6	337	1	YK68_YEAST	P36164 saccharomyc
11	7	8.6	352	1	CITC_ECOLI	P77390 escherichia
12	7	8.6	368	1	ROX1_YEAST	P25042 saccharomyc
13	7	8.6	591	1	ALU1_HUMAN	P39188 homo sapien
14	7	8.6	603	1	ALU4_HUMAN	P39191 homo sapien
15	7	8.6	804	1	SYL_BACSU	P36430 bacillus su
16	7	8.6	806	1	SYL_BACHD	Q9k7s8 bacillus ha
17	7	8.6	889	1	C122_HUMAN	Q9bzq6 homo sapien
18	7	8.6	1230	1	UGS4_SOLTU	Q43846 solanum tub
19	6	7.4	66	1	HSP1_ISOMA	P42136 isoodon mac
20	6	7.4	68	1	HSP1_PERGU	P42147 perameles g
21	6	7.4	93	1	HIS2_MYCTU	O33257 mycobacteri
22	6	7.4	118	1	RS20_SCHPO	O74893 schizosacch
23	6	7.4	134	1	SY21_HUMAN	O00585 homo sapien
24	6	7.4	140	1	RL23_DROME	P48159 drosophila
25	6	7.4	153	1	IF1A_YEAST	P38912 saccharomyc
26	6	7.4	154	1	YK01_AERPE	Q9yad8 aeropyrum p
27	6	7.4	160	1	HBL1_ARATH	O24520 arabidopsis
28	6	7.4	168	1	RR7_CHLRE	P48267 chlamydomon
29	6	7.4	174	1	IL1X_BOVIN	O77482 bos taurus
30	6	7.4	176	1	DPSA_SYNP7	Q55024 synechococc
31	6	7.4	189	1	COAT_TYMV	P03608 turnip yell
32	6	7.4	189	1	COAT_TYMVA	P20125 turnip yell
33	6	7.4	193	1	KITH_HAEIN	P44309 haemophilus
34	6	7.4	201	1	COAE_BACHD	Q9k857 bacillus ha
35	6	7.4	206	1	COX3_BACFI	Q04442 bacillus fi
36	6	7.4	211	1	CLDD_MOUSE	Q9z0s4 mus musculu
37	6	7.4	214	1	GRP2_NICSY	P27484 nicotiana s
38	6	7.4	231	1	TRMD_MYCGE	P47683 mycoplasma
39	6	7.4	252	1	THI4_PYRAB	Q9v0j8 pyrococcus
40	6	7.4	255	1	THI4_PYRHO	O59082 pyrococcus
41	6	7.4	262	1	TRPA_AQUAE	O67502 aquifex aeo
42	6	7.4	281	1	RP32_HAEIN	P44404 haemophilus
43	6	7.4	284	1	RP32_CITFR	P11539 citrobacter
44	6	7.4	284	1	RP32_ECOLI	P00580 escherichia
45	6	7.4	285	1	RP32_ENTCL	P50508 enterobacte

ALIGNMENTS

RESULT 1
RPR_DROME
ID RPR_DROME STANDARD; PRT; 65 AA.
AC Q24475; Q9VVP7;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Reaper protein.
GN RPR OR CG4319.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R; TISSUE=Eye imaginal disk;
RX MEDLINE=94225205; PubMed=8171319;
RA White K., Grether M.E., Abrams J.M., Young L., Farrell K., Steller H.;
RT "Genetic control of programmed cell death in Drosophila.";
RL Science 264:677-683(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

```
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: PLAYS A CENTRAL AND GLOBAL REGULATORY FUNCTION FOR THE
CC       INITIATION OF APOPTOSIS. ECTOPIC EXPRESSION IN THE DEVELOPING EYE
CC       RESULTS IN A SMALL EYE OWING TO EXCESS CELL DEATH.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSION CORRESPONDS TO THE PATTERN OF
CC       PROGRAMMED CELL DEATH IN THE EMBRYO.
CC   -!- SIMILARITY: LIMITED AT THE N-TERMINAL, TO HID AND GRIM.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L31631; AAA18983.1; -.
DR   EMBL; AE003520; AAF49264.1; -.
DR   FlyBase; FBgn0011706; rpr.
KW   Apoptosis.
SQ   SEQUENCE   65 AA;  7682 MW;  57F231379AFEEA3C CRC64;

  Query Match             9.9%;  Score 8;  DB 1;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 0.17;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         35 QKEQQILR 42
              ||||||||
Db         19 QKEQQILR 26


RESULT 2
SUHA_HUMAN
ID   SUHA_HUMAN     STANDARD;      PRT;   284 AA.
AC   Q06520;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Alcohol sulfotransferase (EC 2.8.2.2) (Hydroxysteroid
DE   Sulfotransferase) (HST) (Dehydroepiandrosterone sulfotransferase)
DE   (DHEA-ST) (ST2) (ST2A3).
GN   SULT2A1 OR STD OR HST.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 60-64; 104-119 AND 273-284.
RC   TISSUE=Liver;
RX   MEDLINE=93143674; PubMed=7678732;
RA   Comer K.A., Falany J.L., Falany C.N.;
RT   "Cloning and expression of human liver dehydroepiandrosterone
RT   sulphotransferase.";
RL   Biochem. J. 289:233-240(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 80-107 AND 176-198.
RC   TISSUE=Liver;
RX   MEDLINE=92269778; PubMed=1588921;
RA   Otterness D.M., Wieben E.D., Wood T.C., Watson R.W.G., Madden B.J.,
RA   McCormick D.J., Weinshilboum R.M.;
RT   "Human liver dehydroepiandrosterone sulfotransferase: molecular
RT   cloning and expression of cDNA.";
RL   Mol. Pharmacol. 41:865-872(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=96034512; PubMed=7589785;
RA   Forbes K.J., Hagen M., Coughtrie M.W.H., Glatt H.R., Hume R.;
RT   "Human fetal adrenal hydroxysteroid sulphotransferase: cDNA cloning,
RT   stable expression in V79 cells and functional characterisation of the
RT   expressed enzyme.";
RL   Mol. Cell. Endocrinol. 112:53-60(1995).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95322029; PubMed=7598806;
RA   Luu-The V., Dufort I., Paquet N., Reimnitz G., Labrie F.;
RT   "Structural characterization and expression of the human
RT   dehydroepiandrosterone sulfotransferase gene.";
RL   DNA Cell Biol. 14:511-518(1995).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95225980; PubMed=7710689;
RA   Otterness D.M., Her C., Aksoy S., Kimura S., Wieben E.D.,
RA   Weinshilboum R.M.;
RT   "Human dehydroepiandrosterone sulfotransferase gene: molecular
RT   cloning and structural characterization.";
RL   DNA Cell Biol. 14:331-341(1995).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92392364; PubMed=1520333;
RA   Kong A.-N.T., Yang L., Ma M., Tao D., Bjornsson T.D.;
RT   "Molecular cloning of the alcohol/hydroxysteroid form (hSTa) of
RT   sulfotransferase from human liver.";
RL   Biochem. Biophys. Res. Commun. 187:448-454(1992).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES THE SULFATION OF STEROIDS AND BILE ACIDS IN
CC       THE LIVER AND ADRENAL GLANDS.
CC   -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + an alcohol =
CC       adenosine 3',5'-bisphosphate + an alkyl sulfate.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: LIVER, ADRENAL AND AT LOWER LEVEL IN THE
CC       KIDNEY. IS PRESENT IN HUMAN FETUS IN HIGHER LEVEL IN THE ADRENAL
CC       THAN THE LIVER AND THE KIDNEY.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- MISCELLANEOUS: ESTROGENS PRESENT IN MATERNAL CIRCULATION IS
CC       PREDOMINANTLY DERIVED FROM FETAL DEHYDROEPIANDOSTERONE SULFATE
CC       WHICH IS HYDROLYZED AND METABOLIZED TO ESTROGENS IN PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L20000; AAA35758.1; -.
DR   EMBL; X70222; CAA49755.1; -.
DR   EMBL; U08024; AAA17749.1; -.
DR   EMBL; U08025; AAA17750.1; -.
DR   EMBL; X84816; CAA59274.1; -.
DR   EMBL; L36196; AAA75491.1; -.
DR   EMBL; L36191; AAA75491.1; JOINED.
DR   EMBL; L36192; AAA75491.1; JOINED.
DR   EMBL; L36193; AAA75491.1; JOINED.
DR   EMBL; L36194; AAA75491.1; JOINED.
DR   EMBL; L36195; AAA75491.1; JOINED.
DR   EMBL; U13061; AAC51353.1; -.
DR   EMBL; U13056; AAC51353.1; JOINED.
DR   EMBL; U13057; AAC51353.1; JOINED.
DR   EMBL; U13058; AAC51353.1; JOINED.
DR   EMBL; U13059; AAC51353.1; JOINED.
DR   EMBL; U13060; AAC51353.1; JOINED.
```

```
DR   EMBL; S43859; AAB23169.2; -.
DR   EMBL; BC020755; AAH20755.1; -.
DR   HSSP; P50224; 1CJM.
DR   Genew; HGNC:11458; SULT2A1.
DR   MIM; 125263; -.
DR   InterPro; IPR000863; Sulfotransferase.
DR   Pfam; PF00685; Sulfotransfer; 1.
DR   ProDom; PD001218; Sulfotransferase; 1.
KW   Transferase; Steroid metabolism.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   BINDING     248    254       PAPS BINDING SITE (POTENTIAL).
FT   CONFLICT     62     62       A -> P (IN REF. 1; AA SEQUENCE).
FT   CONFLICT     89     89       T -> S (IN REF. 1).
FT   CONFLICT    118    118       L -> D (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    158    158       L -> V (IN REF. 6).
SQ   SEQUENCE   284 AA;  33648 MW;  3C89C7597833EBA1 CRC64;

  Query Match             9.9%;  Score 8;  DB 1;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 0.59;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 FRSETLRK 55
              ||||||||
Db         17 FRSETLRK 24


RESULT 3
SUHA_MACFA
ID   SUHA_MACFA     STANDARD;      PRT;   284 AA.
AC   P52842;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Alcohol sulfotransferase (EC 2.8.2.2) (Hydroxysteroid
DE   sulfotransferase) (HST).
GN   SULT2A1 OR STD.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Ogura K., Satsukawa M., Kato K., Okuda H., Watabe T.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYSES THE SULFATION OF STEROIDS AND BILE ACIDS IN
CC       THE LIVER AND ADRENAL GLANDS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + an alcohol =
CC       adenosine 3',5'-bisphosphate + an alkyl sulfate.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D85521; BAA12823.1; -.
DR   HSSP; P49891; 1AQU.
DR   InterPro; IPR000863; Sulfotransferase.
DR   Pfam; PF00685; Sulfotransfer; 1.
DR   ProDom; PD001218; Sulfotransferase; 1.
KW   Transferase; Steroid metabolism.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   BINDING     248    254       PAPS BINDING SITE (POTENTIAL).
SQ   SEQUENCE   284 AA;  33789 MW;  8A8DC56BC0B7A98D CRC64;

  Query Match             9.9%;  Score 8;  DB 1;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 0.59;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         48 FRSETLRK 55
              ||||||||
Db         17 FRSETLRK 24


RESULT 4
ALU2_HUMAN
ID   ALU2_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39189;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Alu subfamily SB sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95021758; PubMed=7935834;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation
RT   closely connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC       THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC       CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC       FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC       REPEATS.
CC   -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC       CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC   -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC       POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC       ACID SEQUENCES.
CC   -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC       PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC       ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC       ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC       CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC       LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC       ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC       GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC       CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC       OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC       CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC       DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC       THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC       BEING REPORTED.
CC   -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC       WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC       PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
```

```
CC       CODING NUCLEOTIDE SEQUENCE.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U14568; -; NOT_ANNOTATED_CDS.
KW   Hypothetical protein.
FT   DOMAIN       1     96       FRAME-1.
FT   DOMAIN      100    194      FRAME-2.
FT   DOMAIN      198    292      FRAME-3.
FT   DOMAIN      296    391      FRAME-4.
FT   DOMAIN      395    489      FRAME-5.
FT   DOMAIN      493    587      FRAME-6.
SQ   SEQUENCE   587 AA;  63703 MW;  3EAAB3E3E3929203 CRC64;

  Query Match             9.9%;  Score 8;  DB 1;  Length 587;
  Best Local Similarity   100.0%;  Pred. No. 1.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 SLGDRARL 26
              ||||||||
Db        183 SLGDRARL 190


RESULT 5
ALU3_HUMAN
ID   ALU3_HUMAN     STANDARD;      PRT;   587 AA.
AC   P39190;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Alu subfamily SB1 sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95021758; PubMed=7935834;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation
RT   closely connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC       THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC       CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC       FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC       REPEATS.
CC   -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC       CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC   -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC       POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC       ACID SEQUENCES.
CC   -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC       PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC       ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC       ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC       CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC       LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC       ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC       GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC       CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC       OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC       CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC       DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC       THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC       BEING REPORTED.
CC   -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC       WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC       PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC       CODING NUCLEOTIDE SEQUENCE.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U14569; -; NOT_ANNOTATED_CDS.
KW   Hypothetical protein.
FT   DOMAIN       1     96       FRAME-1.
FT   DOMAIN      100    194      FRAME-2.
FT   DOMAIN      198    292      FRAME-3.
FT   DOMAIN      296    391      FRAME-4.
FT   DOMAIN      395    489      FRAME-5.
FT   DOMAIN      493    587      FRAME-6.
SQ   SEQUENCE   587 AA;  63573 MW;  85C4155726DEF235 CRC64;

  Query Match             9.9%;  Score 8;  DB 1;  Length 587;
  Best Local Similarity   100.0%;  Pred. No. 1.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 SLGDRARL 26
              ||||||||
Db        183 SLGDRARL 190


RESULT 6
ALU8_HUMAN
ID   ALU8_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39195;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Alu subfamily SX sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95021758; PubMed=7935834;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
```

```
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation
RT   closely connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC       THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC       CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC       FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC       REPEATS.
CC   -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC       CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC   -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC       POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC       ACID SEQUENCES.
CC   -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC       PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC       ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC       ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC       CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC       LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC       ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC       GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC       CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC       OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC       CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC       DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC       THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC       BEING REPORTED.
CC   -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC       WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC       PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC       CODING NUCLEOTIDE SEQUENCE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U14574; -; NOT_ANNOTATED_CDS.
KW   Hypothetical protein.
FT   DOMAIN       1     96       FRAME-1.
FT   DOMAIN       100   195      FRAME-2.
FT   DOMAIN       199   294      FRAME-3.
FT   DOMAIN       298   393      FRAME-4.
FT   DOMAIN       397   492      FRAME-5.
FT   DOMAIN       496   591      FRAME-6.
SQ   SEQUENCE   591 AA;  64395 MW;  AC8154AD8A6BB280 CRC64;

  Query Match             9.9%;  Score 8;  DB 1;  Length 591;
  Best Local Similarity   100.0%;  Pred. No. 1.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 SLGDRARL 26
              ||||||||
Db         85 SLGDRARL 92


RESULT 7
FLHD_XENNE
ID   FLHD_XENNE     STANDARD;      PRT;   116 AA.
AC   Q9X9F2;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Flagellar transcriptional activator flhD.
GN   FLHD.
OS   Xenorhabdus nematophilus.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Xenorhabdus.
OX   NCBI_TaxID=628;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F1;
RA   Givaudan A.G., Lanois A.;
RT   "flhDC gene disruptions leads to pleiotropic phenotypes.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Transcriptional activator. Together with flhC it acts as
CC       a compound sigma factor that activates class 2 flagellar genes (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE FLHD FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ012828; CAB41407.1; -.
KW   Transcription; Transcription regulation; Sigma factor;
KW   DNA-directed RNA polymerase; DNA-binding; Flagella.
SQ   SEQUENCE   116 AA;  13303 MW;  89DBFF8175532828 CRC64;

  Query Match             8.6%;  Score 7;  DB 1;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LLSTHLF 8
              |||||||
Db         95 LLSTHLF 101


RESULT 8
INVO_CANFA
ID   INVO_CANFA     STANDARD;      PRT;   285 AA.
AC   P18174;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Involucrin.
GN   IVL.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90348475; PubMed=2385171;
RA   Tseng H., Green H.;
RT   "The involucrin genes of pig and dog: comparison of their segments of
RT   repeats with those of prosimians and higher primates.";
RL   Mol. Biol. Evol. 7:293-302(1990).
CC   -!- FUNCTION: Part of the insoluble cornified cell envelope (CE) of
CC       stratified squamous epithelia.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic. Constituent of the scaffolding
CC       of the cornified envelope.
CC   -!- TISSUE SPECIFICITY: Keratinocytes of epidermis and other
CC       stratified squamous epithelia.
CC   -!- PTM: Substrate of transglutaminase. Specific glutamines or lysines
CC       are cross-linked to keratins, desmoplakin and to inter involucrin
CC       molecules.
```

```
CC   -!- SIMILARITY: BELONGS TO THE INVOLUCRIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34442; AAA30853.1; -.
DR   InterPro; IPR002360; Involucrin.
DR   PROSITE; PS00795; INVOLUCRIN; 1.
KW   Keratinization; Repeat.
SQ   SEQUENCE   285 AA;  33384 MW;  DCE1BD88B9248BEA CRC64;

  Query Match             8.6%;  Score 7;  DB 1;  Length 285;
  Best Local Similarity   100.0%;  Pred. No. 6.5;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         32 KQQQKEQ 38
              |||||||
Db        198 KQQQKEQ 204


RESULT 9
Y687_HAEIN
ID   Y687_HAEIN     STANDARD;      PRT;   304 AA.
AC   P71356;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical transport protein HI0687.
GN   HI0687.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: BELONGS TO THE EAMA TRANSPORTER FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32751; AAC22347.1; -.
DR   TIGR; HI0687; -.
DR   InterPro; IPR000620; DUF6.
DR   Pfam; PF00892; DUF6; 2.
KW   Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT   TRANSMEM      9     29       POTENTIAL.
FT   TRANSMEM     67     87       POTENTIAL.
FT   TRANSMEM    100    120       POTENTIAL.
FT   TRANSMEM    131    151       POTENTIAL.
FT   TRANSMEM    159    179       POTENTIAL.
FT   TRANSMEM    189    209       POTENTIAL.
FT   TRANSMEM    222    242       POTENTIAL.
FT   TRANSMEM    252    272       POTENTIAL.
FT   TRANSMEM    278    298       POTENTIAL.
SQ   SEQUENCE   304 AA;  33887 MW;  CC7095529E8E4FB3 CRC64;

  Query Match             8.6%;  Score 7;  DB 1;  Length 304;
  Best Local Similarity   100.0%;  Pred. No. 6.9;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 LFIYLFI 13
              |||||||
Db         71 LFIYLFI 77


RESULT 10
YK68_YEAST
ID   YK68_YEAST     STANDARD;      PRT;   337 AA.
AC   P36164;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   Hypothetical 38.3 kDa protein in PRP16-SRP40 intergenic region.
GN   YKR088C OR YKR408.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94262327; PubMed=8203164;
RA   Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
RA   Remacha M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;
RT   "The complete sequence of an 18,002 bp segment of Saccharomyces
RT   cerevisiae chromosome XI contains the HBS1, MRP-L20 and PRP16 genes,
RT   and six new open reading frames.";
RL   Yeast 10:231-245(1994).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z27116; CAA81639.1; -.
DR   EMBL; Z28313; CAA82167.1; -.
DR   PIR; S38166; S38166.
DR   PIR; S39129; S39129.
DR   SGD; S0001796; YKR088C.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     96    115       POTENTIAL.
FT   TRANSMEM    138    162       POTENTIAL.
FT   TRANSMEM    173    191       POTENTIAL.
FT   TRANSMEM    222    246       POTENTIAL.
FT   TRANSMEM    253    271       POTENTIAL.
FT   TRANSMEM    287    309       POTENTIAL.
SQ   SEQUENCE   337 AA;  38311 MW;  7EA95DD4E5AF77FE CRC64;

  Query Match             8.6%;  Score 7;  DB 1;  Length 337;
  Best Local Similarity   100.0%;  Pred. No. 7.5;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 LFIYLFI 13
              |||||||
Db        264 LFIYLFI 270


RESULT 11
```

```
CITC_ECOLI
ID   CITC_ECOLI     STANDARD;      PRT;   352 AA.
AC   P77390; O54337; Q9R2T4;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   [Citrate [pro-3S]-lyase] ligase (EC 6.2.1.22) (Citrate lyase
DE   synthetase) (Acetate:SH-citrate lyase ligase).
GN   CITC OR B0618.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA   Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA   Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RA   Ingmer H., Cohen S.N.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ACETYLATION OF PROSTHETIC GROUP (2-(5''-PHOSPHORIBOSYL)-
CC       3'-DEPHOSPHOCOENZYME-A) OF THE GAMMA SUBUNIT OF CITRATE LYASE.
CC   -!- CATALYTIC ACTIVITY: ATP + acetate + [citrate (pro-3S)-lyase]
CC       (thiol form) = AMP + diphosphate + [citrate (pro-3S)-lyase]
CC       (acetyl form).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000166; AAC73719.1; ALT_INIT.
DR   EMBL; U82598; AAB40818.1; ALT_INIT.
DR   EMBL; D90702; BAA35254.1; ALT_INIT.
DR   EMBL; D90703; BAA35260.1; ALT_INIT.
DR   EMBL; U46667; AAC28950.1; -.
DR   EcoGene; EG13645; citC.
DR   InterPro; IPR005216; Cit_ly_ligase.
DR   InterPro; IPR004821; Cyt_tran_rel.
DR   InterPro; IPR000182; GCN5acetyltransf.
DR   Pfam; PF00583; Acetyltransf; 1.
DR   TIGRFAMs; TIGR00124; cit_ly_ligase; 1.
DR   TIGRFAMs; TIGR00125; cyt_tran_rel; 1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   352 AA;  40077 MW;  F5894FFCD0F06518 CRC64;

  Query Match             8.6%;  Score 7;  DB 1;  Length 352;
  Best Local Similarity   100.0%;  Pred. No. 7.8;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 STHLFIY 10
              |||||||
Db         90 STHLFIY 96


RESULT 12
ROX1_YEAST
ID   ROX1_YEAST     STANDARD;      PRT;   368 AA.
AC   P25042;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ROX1 repressor (Hypoxic function repressor) (Heme-dependent repression
DE   factor).
GN   ROX1 OR YPR065W OR YP9499.20.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94019282; PubMed=8413209;
RA   Balasubramanian B., Lowry C.V., Zitomer R.S.;
RT   "The Rox1 repressor of the Saccharomyces cerevisiae hypoxic genes is
RT   a specific DNA-binding protein with a high-mobility-group motif.";
RL   Mol. Cell. Biol. 13:6071-6078(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313271; PubMed=9169875;
RA   Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA   Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA   Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA   Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA   Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA   Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA   Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
RA   Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA   Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA   Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA   Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA   Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA   Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA   Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA   Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA   Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL   Nature 387:103-105(1997).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=96174644; PubMed=8600445;
RA   di Flumeri C., Liston P., Acheson N.H., Keng T.;
RT   "The HMG domain of the ROX1 protein mediates repression of HEM13
RT   through overlapping DNA binding and oligomerization functions.";
RL   Nucleic Acids Res. 24:808-815(1996).
CC   -!- FUNCTION: TRANSCRIPTION FACTOR THAT REPRESSES THE EXPRESSION OF
CC       HEM13, COX5B, ANB1, CYC7 OR AAC3 (HYPOXIC FUNCTION). BINDS TO THE
CC       DNA SEQUENCE 5'-RRRTAACAAGAG-3'.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- INDUCTION: BY HEME.
CC   -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X60458; CAA42991.1; -.
DR   EMBL; Z49219; CAA89182.1; -.
DR   EMBL; Z71255; CAA94973.1; -.
DR   PIR; S17015; S17015.
DR   HSSP; Q05066; 1HRY.
DR   TRANSFAC; T01286; -.
DR   SGD; S0006269; ROX1.
DR   InterPro; IPR000910; HMG_12_box.
DR   Pfam; PF00505; HMG_box; 1.
DR   SMART; SM00398; HMG; 1.
KW   Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT   DNA_BIND     14     83       HMG BOX.
FT   DOMAIN      102    123       GLN-RICH.
SQ   SEQUENCE   368 AA;  41838 MW;  3B27442D7DEE3DBD CRC64;

  Query Match             8.6%;  Score 7;  DB 1;  Length 368;
  Best Local Similarity   100.0%;  Pred. No. 8.1;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         33 QQQKEQQ 39
              |||||||
Db        106 QQQKEQQ 112


RESULT 13
ALU1_HUMAN
ID   ALU1_HUMAN     STANDARD;      PRT;   591 AA.
AC   P39188;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Alu subfamily J sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95021758; PubMed=7935834;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation
RT   closely connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC       THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC       CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC       FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC       REPEATS.
CC   -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC       CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC   -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC       POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC       ACID SEQUENCES.
CC   -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC       PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC       ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC       ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC       CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC       LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC       ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC       GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC       CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC       OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC       CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC       DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC       THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC       BEING REPORTED.
CC   -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC       WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC       PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC       CODING NUCLEOTIDE SEQUENCE.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U14567; -; NOT_ANNOTATED_CDS.
KW   Hypothetical protein.
FT   DOMAIN        1     96       FRAME-1.
FT   DOMAIN      100    195       FRAME-2.
FT   DOMAIN      199    294       FRAME-3.
FT   DOMAIN      298    393       FRAME-4.
FT   DOMAIN      397    492       FRAME-5.
FT   DOMAIN      496    591       FRAME-6.
SQ   SEQUENCE   591 AA;  63790 MW;  665D395735519D95 CRC64;

  Query Match             8.6%;  Score 7;  DB 1;  Length 591;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         19 SLGDRAR 25
              |||||||
Db         85 SLGDRAR 91


RESULT 14
ALU4_HUMAN
ID   ALU4_HUMAN     STANDARD;      PRT;   603 AA.
AC   P39191;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Alu subfamily SB2 sequence contamination warning entry.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95021758; PubMed=7935834;
RA   Claverie J.-M., Makalowski W.;
RT   "Alu alert.";
RL   Nature 371:752-752(1994).
RN   [2]
RP   CONCEPT.
RX   MEDLINE=92241891; PubMed=1572661;
RA   Claverie J.-M.;
RT   "Identifying coding exons by similarity search: alu-derived and other
RT   potentially misleading protein sequences.";
```

```
RL   Genomics 12:838-841(1992).
RN   [3]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=88333009; PubMed=3138422;
RA   Quentin Y.;
RT   "The Alu family developed through successive waves of fixation
RT   closely connected with primate lineage history.";
RL   J. Mol. Evol. 27:194-202(1988).
RN   [4]
RP   ALU FAMILIES CLASSIFICATION.
RX   MEDLINE=91178815; PubMed=1706781;
RA   Jurka J., Milosavljevic A.;
RT   "Reconstruction and analysis of human Alu genes.";
RL   J. Mol. Evol. 32:105-121(1991).
CC   -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC       THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC       CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC       FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC       REPEATS.
CC   -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC       CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC   -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC       POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC       ACID SEQUENCES.
CC   -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC       PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC       ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC       ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC       CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC       LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC       ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC       GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC       CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC       OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC       CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC       DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC       THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC       BEING REPORTED.
CC   -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC       WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC       PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC       CODING NUCLEOTIDE SEQUENCE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U14570; -; NOT_ANNOTATED_CDS.
KW   Hypothetical protein.
FT   DOMAIN        1     98       FRAME-1.
FT   DOMAIN      102    199       FRAME-2.
FT   DOMAIN      202    300       FRAME-3.
FT   DOMAIN      304    401       FRAME-4.
FT   DOMAIN      405    502       FRAME-5.
FT   DOMAIN      506    603       FRAME-6.
SQ   SEQUENCE   603 AA;  65272 MW;  B8AAD0AD46BEA114 CRC64;

  Query Match             8.6%;  Score 7;  DB 1;  Length 603;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         20 LGDRARL 26
              |||||||
Db         88 LGDRARL 94


RESULT 15
SYL_BACSU
ID   SYL_BACSU      STANDARD;      PRT;   804 AA.
AC   P36430; O34465;
DT   01-JUN-1994 (Rel. 29, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN   LEUS.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92283747; PubMed=1317842;
RA   Vander Horn P.B., Zahler S.A.;
RT   "Cloning and nucleotide sequence of the leucyl-tRNA synthetase gene
RT   of Bacillus subtilis.";
RL   J. Bacteriol. 174:3928-3935(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98048467; PubMed=9387221;
RA   Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT   "Sequencing and functional annotation of the Bacillus subtilis genes
RT   in the 200 kb rrnB-dnaB region.";
RL   Microbiology 143:3431-3441(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC       diphosphate + L-leucyl-tRNA(Leu).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M88581; AAA22571.1; -.
```

```
DR   EMBL; AF008220; AAC00259.1; -.
DR   EMBL; Z99119; CAB15010.1; -.
DR   PIR; A41882; A41882.
DR   SubtiList; BG10676; leuS.
DR   InterPro; IPR002302; Leu-tRNAsynt1a.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00985; TRNASYNTHLEU.
DR   TIGRFAMs; TIGR00396; leuS_bact; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
FT   SITE         40     51       "HIGH" REGION.
FT   SITE        576    580       "KMSKS" REGION.
FT   BINDING     579    579       ATP (BY SIMILARITY).
FT   CONFLICT    186    186       P -> L (IN REF. 1).
FT   CONFLICT    195    195       T -> N (IN REF. 1).
FT   CONFLICT    247    281       RPDTLFGATYTVLAPEHALVENITTAEQKEAVEAY -> DQ
FT                                IRCLALHTLSLPRNTHWWKTSQRQSKKKLLKLI (IN
FT                                REF. 1).
SQ   SEQUENCE   804 AA;  91542 MW;  306FD5A98FE5C47E CRC64;

  Query Match             8.6%;  Score 7;  DB 1;  Length 804;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         56 TGKKGRR 62
              |||||||
Db        479 TGKKGRR 485


Search completed: May 29, 2003, 15:31:09
Job time : 23 secs
```